

High-resolution whole-organ mapping with SNPs and its significance to early events of carcinogenesis

Tomasz Tuziak^{1,*}, Joon Jeong^{1,*}, Tadeusz Majewski¹, Mi-Sook Kim¹, Jordan Steinberg², Zhi Wang¹, Dong-Sup Yoon¹, Tang C Kuang³, Keith Baggerly⁴, Dennis Johnston³ and Bogdan Czerniak¹

¹Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA;

²Department of Urology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA;

³Department of Biomathematics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA and ⁴Department of Biostatistics & Applied Mathematics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

We attempted to identify deleted segments in two model tumor suppressor gene loci on chromosomes 13q14 and 17p13 that were associated with clonal expansion of *in situ* bladder preneoplasia using single nucleotide polymorphisms (SNPs)-based whole-organ histologic and genetic mapping. For mapping with SNPs, the sequence-based maps spanning approximately 27 and 5 Mb centered around RB1 and p53, respectively, were assembled. The integrated gene and SNP maps of the regions were used to select 661 and 960 SNPs, which were genotyped by pyrosequencing. Genotyping of SNPs was performed on DNA samples corresponding to histologic maps of the entire bladder mucosa in human cystectomy specimens with invasive urothelial carcinoma. By using this approach, we have identified deleted regions associated with clonal expansion of intraurothelial neoplasia; which ranged from 0.001 to 4.3 Mb (average 0.67 Mb) and formed clusters of discontinuous deleted segments. The high resolution of such maps is a prerequisite for future positional targeting of genes involved in early phases of bladder neoplasia. This approach also permits analysis of the overall genomic landscape of the involved region and discloses that a unique composition of noncoding DNA characterized by a high concentration of repetitive sequences may predispose to deletions.

Laboratory Investigation (2005) 85, 689–701. doi:10.1038/labinvest.3700270

Keywords: single nucleotide polymorphic site; bladder cancer; genetic mapping; clonal expansion of preneoplasia

Cancer develops via multiple, complex steps, many of which precede the development of microscopically recognizable preneoplastic conditions.^{1–3} Recent studies indicate that common human malignancies, including bladder cancer, begin as clonal *in situ* expansion of phenotypically normal cells, which may form large plaques involving the affected mucosa.^{4,5} Identification of chromosomal regions and their target genes involved in the development of such clinically and microscopically occult preneoplastic lesions may provide valuable clues to the incipient events of human carcinogenesis.^{6,7}

Ultimately, the results could lead to the development of novel markers for early cancer detection and prevention.

We have previously reported the identification of several putative tumor suppressor gene loci involved in early preinvasive phases of human bladder cancer.^{8–11} These loci were found by searching for allelic imbalances using DNA hypervariable markers and a whole-organ histologic and genetic mapping strategy developed in our laboratory. The ultimate identification of genes mapping to the deleted regions was hampered by the low resolution of the original recombination-based maps and the absence of continuous genome sequence maps, which were not available in the early phases of the human genome project.

In this paper, we present the strategy that combines whole-organ histologic maps and high-resolution genetic mapping using single nucleotide polymorphic (SNP) sites. It facilitates precise iden-

Correspondence: Dr B Czerniak, MD, PhD, Department of Pathology, Box 085, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX, USA.
E-mail: bczernia@mdanderson.org

*Both these authors have contributed equally to this work.
Received 7 January 2005; revised and accepted 8 February 2005

tification of deleted chromosomal regions and their positional candidate genes associated with growth advantage and clonal expansion of *in situ* preneoplasia.^{12–14} Here, we concentrate on the technological aspects of this approach referred to as high-resolution whole-organ mapping with SNPs and present the pattern of losses associated with clonal expansion of *in situ* bladder preneoplasia in two model tumor suppressor gene loci (RB1 and p53).

Methods

Whole-Organ Histologic and Genetic Mapping

Whole-organ histologic and genetic mapping was performed as described in our previous publications.^{8–11} In brief, each fresh cystectomy specimen, taken from a patient who had bladder cancer, was opened longitudinally along the anterior wall of the bladder and pinned down to a paraffin block. The entire mucosa was divided into 1 × 2 cm rectangular samples and evaluated microscopically on frozen sections. The tissue of interest was microdissected from the frozen block, and cell suspensions were prepared by mechanically scraping the urothelial mucosa or gently shaking invasive tumor samples. Only those specimens that yielded more than 90% microscopically recognizable, intact urothelial, dysplasia of tumor were accepted for the study and used for DNA extraction. The intraurothelial precancerous changes were classified as mild, moderate, and severe dysplasia of carcinoma *in situ*. The tumors were classified according to the three-tier World Health Organization histological grading system and growth pattern (papillary vs nonpapillary).¹⁵ The depth of invasion was recorded according to the TNM (tumor–node–metastasis) staging system.¹⁶ In this report, we present the data using two whole maps of cystectomy specimens, which yield 62 and 40 DNA samples, respectively, and both contained microscopically normal urothelium, preneoplastic condition, *in situ* changes, and invasive cancer. In

both maps, a single focus of grade 3 nonpapillary urothelial carcinoma invaded the muscularis propria and was accompanied by extensive *in situ* precancerous lesions ranging from mild dysplasia to carcinoma *in situ*.

Genotyping with SNPs

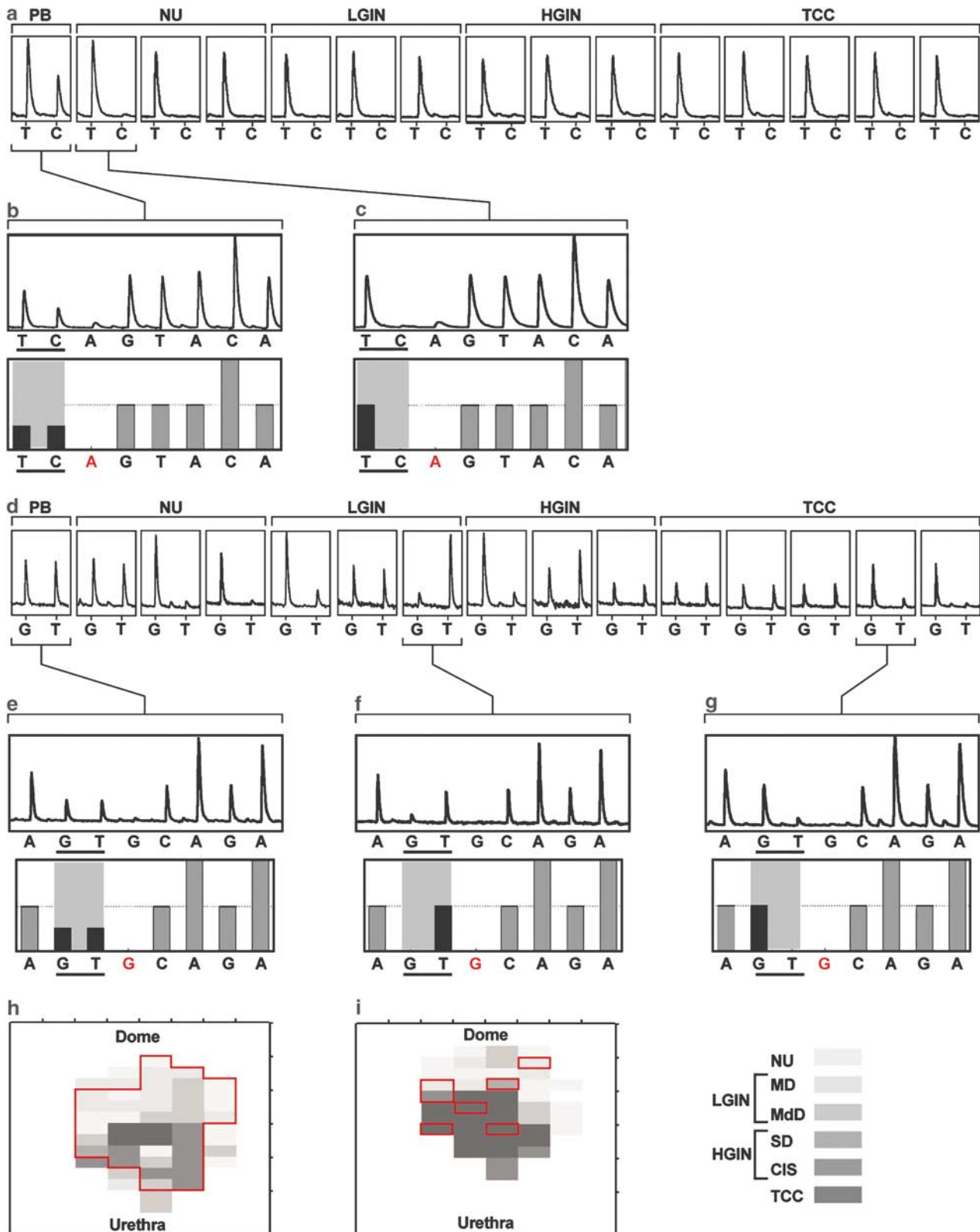
The positions of chromosomal regions associated with an *in situ* clonal expansion mapping to 13q14 and 17p13 were initially identified by the whole-organ histologic and genetic mapping strategy with microsatellite markers. The recombination-based deletional maps of these regions can be obtained at www.mdanderson.org/departments/genomemaps/. For high-resolution mapping with SNPs, the genome sequence maps spanning approximately 27 and 5 Mb, corresponding to microsatellite defined regions in 13q14 and 17p13, respectively, were assembled. The positions of all currently defined SNPs mapping within these regions (approximately 20 000 SNPs per region) were identified in the SNP database (<http://www.ncbi.nlm.nih.gov/SNP>) and integrated with the genome sequence maps. The integrated gene and SNP maps were used to select 661 SNPs mapping within a 27-Mb segment around RB1 and 960 SNPs mapping within a 5-Mb segment around p53.

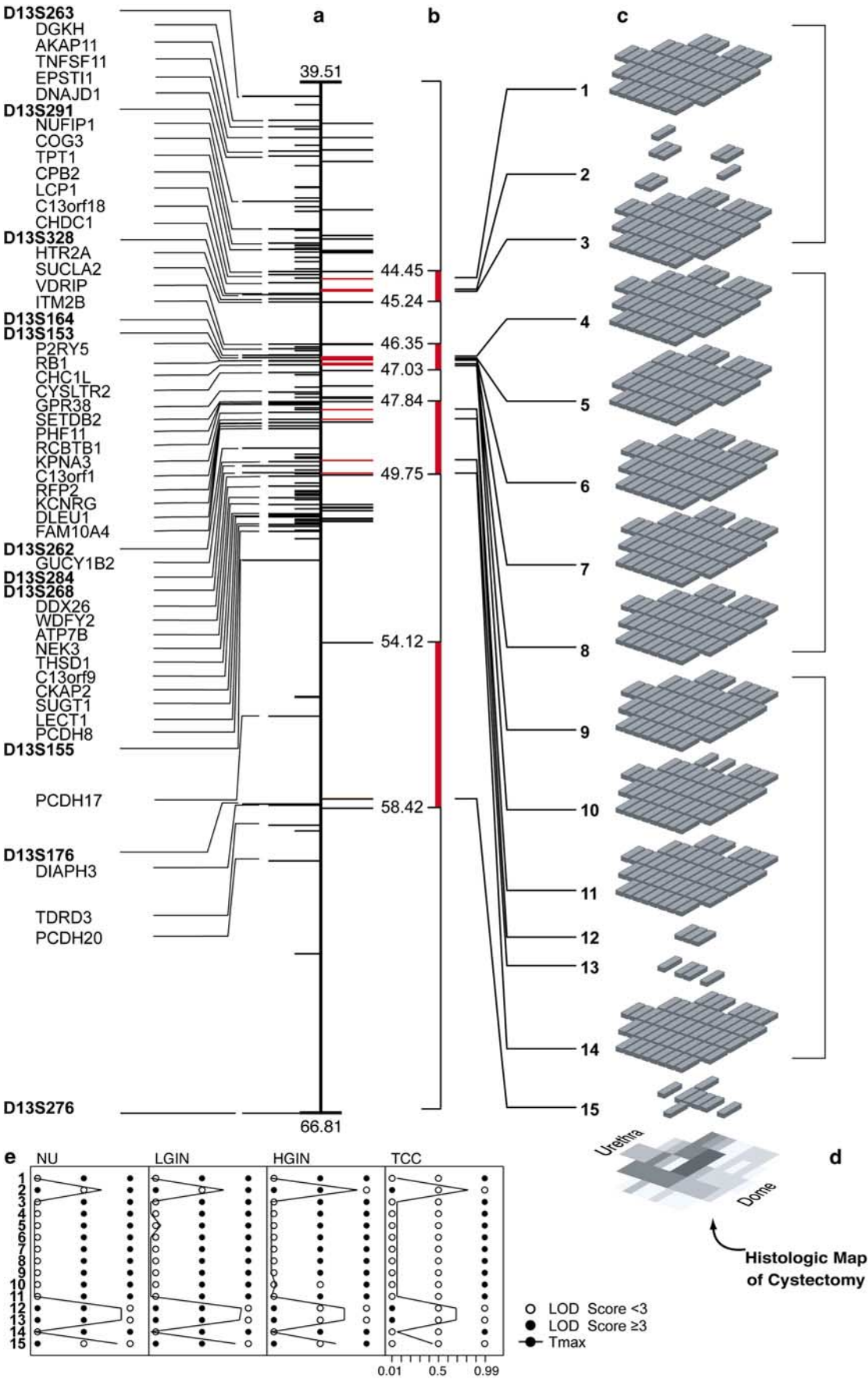
SNPs were genotyped by pyrosequencing methods using an automatic pyrosequencing instrument PSQTM 96 (Pyrosequencing AB, Uppsala, Sweden).^{17,18} In brief, genomic DNA fragments containing SNPs were amplified by PCR with one of each primer biotinylated at the 5'-end. Single-stranded DNA was isolated by streptavidin-coated magnetic beads (Dynabeads M280, Dynal, Oslo, Norway). The sequencing reaction mixture contained a single-stranded DNA with sequencing primer annealed, DNA polymerase, apyrase, luciferase, ATP sulfurylase, adenosine 5'-phosphosulfate, and luciferin. The PCR and sequencing reactions were performed on a TECAN robotic system (GENESIS workstation

Figure 1 Whole-organ histologic mapping, and molecular haplotyping with SNPs. (a–c) An example of molecular haplotyping with SNP showing clonal allelic loss in multiple mucosal samples. (a) Sequence pyrograms of a polymorphic T/C residue showing a clonal loss of C in areas of microscopically normal appearing urothelium, *in situ* preneoplastic conditions (HGIN, LGIN), and invasive carcinoma (TCC). (b, c) Examples of sequence pyrograms with a T/C polymorphism in peripheral blood and a loss of C residue in a mucosal sample from NU are shown in the upper panels. Corresponding histograms showing expected signals of nucleotide incorporation are shown in the lower panels. (d–g) Molecular haplotyping of SNP showing a nonclonal loss of G or T residues in scattered mucosal samples. (d) Sequence pyrograms of SNP showing a loss of T or G residues in scattered mucosal samples. (e–g) Examples of sequence pyrograms with a G/T polymorphism in peripheral blood and random loss of T or G residues in individual mucosal samples are shown in the upper panels. Corresponding histograms showing expected sequences of nucleotide incorporation signals of SNP with retention of both alleles and loss of T or G residues are shown in the lower panels. (h) A histologic map of a cystectomy specimen with an outlined area of mucosa, exhibiting clonal allelic loss of the SNP shown in (a). Note that the area of allelic loss forms a continuous plaque, implying that it was associated with clonal expansion of an *in situ* preneoplasia. (i) A histologic map of cystectomy specimen with outlined areas of mucosa involved by a nonclonal loss of the SNP shown in (d). The polymorphic nucleotide is underlined in sequence pyrograms and their corresponding histograms. A negative control after exposure to the substrates of irrelevant nucleotides follows the polymorphic site and is printed in red. PB, peripheral blood; NU, normal urothelium; LGIN, low-grade intraurothelial neoplasia; HGIN, high-grade intraurothelial neoplasia; TCC, invasive transitional cell carcinoma. Histologic map code: NU, normal urothelium; MD, mild dysplasia; Mdd, moderate dysplasia; SD, severe dysplasia; CIS, carcinoma *in situ*; TCC invasive transitional cell carcinoma. For the purpose of statistical analysis, intraurothelial precursor conditions were classified into two groups: low-grade intraurothelial neoplasia (mild to moderate dysplasia, LGIN) and high-grade intraurothelial neoplasia (severe dysplasia and carcinoma *in situ*, HGIN).

150). The sequence was determined from the measured signal output of light upon nucleotide incorporation. The resulting peaks were analyzed

using pyrosequencing software. A minimum 50% signal intensity reduction from one of the polymorphic nucleotides as compared to the nontumor





DNA of the same individual was used to identify a haplotype defining allelic loss. The distributions of losses identified in individual mucosal samples for each SNP were subsequently superimposed over the histologic maps of the entire cystectomy specimen.

Analysis of SNP-Based Haplotype Distribution Patterns

The distribution patterns of allelic loss defined by SNP-based haplotypes in relation to the development and progression of *in situ* neoplasia were analyzed in superimposed whole-organ histologic and genetic maps by nearest neighbor and binomial maximum likelihood algorithms as previously described in detail.¹¹ In brief, for statistical analyses, the data were organized in 2×4 contingency tables (f_{ji} , $j = 1, 2$; $i = 1, \dots, 4$). The columns designated whether the marker retained or lost heterozygosity and the four grades represented the microscopic status of the urothelium (D_0 , normal urothelium, D_1 , low-grade intraurothelial neoplasia, D_2 , high-grade intraurothelial neoplasia, and D_3 , transitional cell carcinoma (TCC)). For statistical analysis, the intraurothelial precancerous changes were classified into two major groups: low-grade intraurothelial neoplasia (mild and moderate dysplasia, LGIN) and high-grade intraurothelial neoplasia (severe dysplasia and carcinoma *in situ*, HGIN).¹

A nearest-neighbor analysis was performed on the three-dimensional stacks of maps consisting of plots with results of sequencing data for individual SNPs by location on the histologic bladder and genome sequence maps.¹⁹ A region with allelic loss was considered a neighbor of another region of allelic loss if the two were side by side in the same SNP plot or above and below each other. A region with allelic loss was also considered to be connected to another region with allelic loss if there was a continuous string of regions with allelic losses between them. Since the bladder was laid open

and pinned flat, the left-most and right-most regions were also neighbors.

The binomial maximum likelihood analysis was performed in the same 2×4 contingency tables. The maximum likelihood of the binomial distribution was calculated for each SNP and urothelial status (D_i , $i = 0, \dots, 3$) corresponding to normal urothelium (NU), LGIN, HGIN, and TCC. The maximum likelihood for the binomial distribution was used to determine whether a row of data was consistent with a hypothesis of an unchanged (retention of polymorphism) SNP by calculating the log likelihood with

$$l_i = \ln \left(\frac{\theta_i^{f_{1i}} (1 - \theta_i)^{f_{2i}}}{\hat{\theta}_i^{f_{1i}} (1 - \hat{\theta}_i)^{f_{2i}}} \right)$$

for the null hypothesis H_0 , where $\theta = \theta_i$, and $\theta = f_{1i}/(f_{1i} + f_{2i})$ is the maximum likelihood estimate of SNP with retention of polymorphism. Two times the negative of the log likelihood, $-2l_i$, is asymptotically χ^2 with 1 degree of freedom, $\chi^2(1)$. This expression can be written as

$$\begin{aligned} -2l_i &= 2 \ln(10) \log_{10} \left(\frac{\hat{\theta}_i^{f_{1i}} (1 - \hat{\theta}_i)^{f_{2i}}}{\theta_i^{f_{1i}} (1 - \theta_i)^{f_{2i}}} \right) \\ &= 2 \ln(10) \text{LOD}(\hat{\theta}_i; \theta_i) \end{aligned}$$

where $\text{LOD}(\hat{\theta}_i; \theta_i)$ is the LOD-score function evaluated at θ_i .²⁰ Each row of the table for which $-2l_i$ has approximate $\chi^2(1)$ can be tested separately (stringency level 1) or all rows for diagnosis D_i and more advanced (D_i, \dots, D_3) can be combined (stringency level 2) to get

$$f_{1+} = \sum_{j=1}^3 f_{1j}$$

$$f_{2+} = \sum_{j=1}^3 f_{2j}$$

Figure 2 Whole-organ mapping based on allelotyping of SNPs and assembly of three-dimensional display of allelic losses on chromosome 13q14 in a single cystectomy specimen. (a) Genome sequence map of a 27.3-Mb region on chromosome 13q14 (www.mdanderson.org/departments/genomemaps/). The positions of hypervariable DNA markers and positions of the known or predicted genes are depicted by the bars on the left-hand side of the map. The bars on the right-hand side of the map designate the positions of all informative (polymorphic) SNPs identified in peripheral blood of the same patient. Monomorphic (noninformative) SNPs are not shown. (b) Map of deleted segments depicted by red bars. Their predicted borders are defined by the nearest flanking SNPs that retained polymorphism in all mucosal samples (see Figure 4). The positions of SNPs with allelic loss are shown on the right-hand side, and their distribution patterns in bladder mucosa are depicted by the shaded blocks in c. (c) Three-dimensional display of allelic losses depicted by shaded blocks in relation to the distribution of TCC and their association with *in situ* preneoplastic lesions displayed in the whole-organ histologic map of a cystectomy specimen at the bottom of the diagram in (d). The SNPs with loss of polymorphism are numbered consecutively 1–15. (d) Graphic representation of histologic map of cystectomy specimen with the distribution of *in situ* preneoplastic lesions that progressed to invasive TCC. Histologic map code is identical to the code shown in Figure 1. (e) Summary of binomial maximum likelihood analysis testing the relationship between loss of SNP polymorphism and progression of urothelial neoplasia from *in situ* precursor conditions to invasive TCC. Cumulative LOD scores for SNPs with loss of polymorphism were calculated at variable $\theta = (0.01, 0.5, \text{ and } 0.99)$ and tested against T_{\max} . The significance of clonal losses was analyzed for normal urothelium (NU), low-grade intraurothelial neoplasia (LGIN), high-grade intraurothelial neoplasia (HGIN), and TCC. To simplify the data, stringency 1 calculations are presented only. The patterns of significant LOD scores are as described in Methods. The SNPs with loss of polymorphism are numbered consecutively from 1 to 15 as depicted in (c). Note the significant pattern of LOD scores typically parallel to the high T_{\max} values (\circ , LOD score < 3 ; \bullet , LOD score ≥ 3). A cluster of three discontinuous segments of allelic loss associated with clonal expansion, which form overlapping large plaques, not only involving microscopically recognizable *in situ* preneoplastic lesions but also extending into areas of microscopically normal urothelium of bladder mucosa, was identified. The segments were defined by clusters of neighboring SNP with clonal loss of polymorphism as shown by brackets on the right-hand side of the three-dimensional diagram.

$$\text{LOD}(\hat{\theta}_{i+} : \theta_{i+}) = \log_{10} \left(\frac{\hat{\theta}_{i+}^{f_{i+}} (1 - \hat{\theta}_{i+})^{f_{2i+}}}{\theta_{i+}^{f_{i+}} (1 - \theta_{i+})^{f_{2i+}}} \right)$$

which is also $\chi^2(1)$ after adjustment by $2 \ln(10) = 4.605$. Because the maximum likelihood estimates for the individual rows are usually different from each other, the sum of the LOD scores and the sum of the χ^2 's were greater than the combined statistics. A χ^2 test for heterogeneity was appropriate to test the combined estimate. Usually $\theta_i = 0.5$ is used to test linkage in familial disorders with meiotic segregation of the phenotype.²⁰ In reference to sporadic cancer and especially when populations of tested cells represent sequential stages of the process with mitotic transmission of the phenotype, the null hypothesis is more appropriately verified at θ differing from 0.5. For example, a value of 0.99 is more appropriate if the SNP retains polymorphism in the tissue, and a value of 0.01 is more appropriate for determining whether the SNP has lost polymorphism in the later stages of the process, that is, invasive carcinoma. The pattern of LOD scores ≥ 3 at $\theta = 0.01$ or 0.99 and LOD scores < 3 at $\theta = 0.5$ for the same SNP was significant. The strongest association between an SNP with loss of polymorphism and neoplasia is when a LOD score is ≥ 3 at $\theta = 0.99$ and 0.5 and < 3 at $\theta = 0.01$. The use of binomial likelihood analysis and LOD scores in our study is not the same as that commonly used in linkage of familial genetic predispositions for diseases. Here it is intended to be used in its generic mathematical sense as likelihood tests of events.²⁰

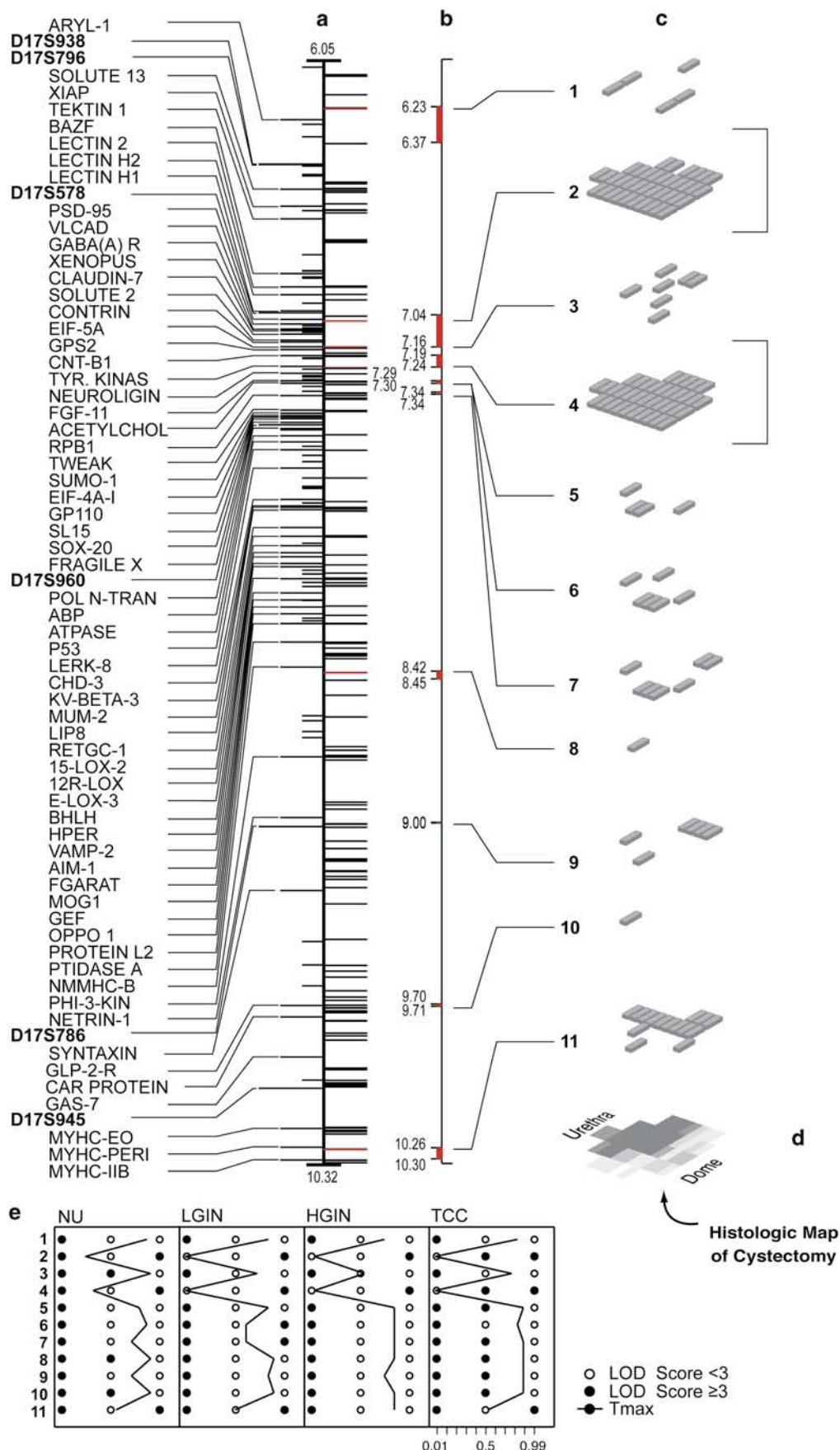
Analysis of Genome Sequences

We analyzed the genome sequences spanning the 27 and 4 Mb around RB1 and p53, respectively.

The genomic maps contained the positions of all known and predicted genes, SNPs, and non-LTR repeats integrated with the positions of deleted regions defined by SNP-based haplotypes from superimposed whole-organ histologic and genetic maps. We concentrated on those deleted regions that showed association with an *in situ* clonal expansion of preneoplastic urothelial cells and analyzed their CpG island and gene contents as well as the frequency of non-LTR repeats.

First, visual investigations of genomic maps and deleted regions suggested that the segments of deletions associated with *in situ* clonal expansions may contain a high density of non-LTR repeats. In order to verify this observation, we started with the null hypothesis that location of the deletions was independent of the repeat sequences density. We tested this hypothesis by analyzing the densities of repeats at fixed width windows centered at the informative SNPs. The repeat sequences density was defined as the number of repeat sequences base pairs in a fixed width window. The width of this window was proportional to the spacing between informative SNPs. The median distance between informative SNPs of approximately 27 kB suggested an optimal window of 13.5 kB on either side of the informative SNP. In the second step, we ranked all repeat sequences according to the number of base pairs and Wilcoxon's test was used to analyze the difference in density of repeats around SNPs showing loss associated with plaque-like clonal expansion, SNPs showing scattered nonclonal pattern of loss and with informative SNPs showing retention of polymorphism. Similar analyses were performed to test the relationship between densities of CpG islands, gene content, and the pattern of allelic loss defined by SNPs (data not shown).

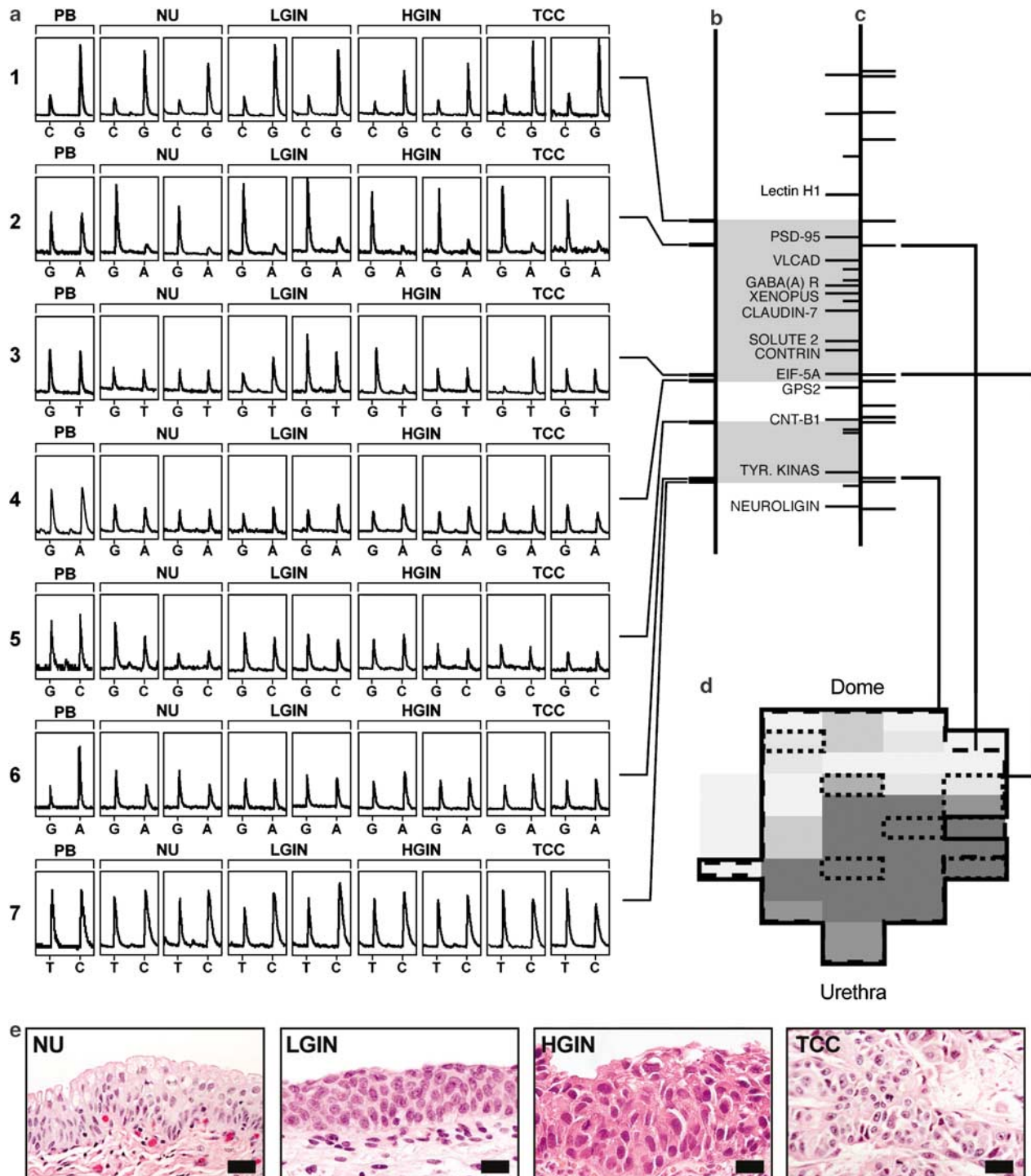
Figure 3 Whole-organ mapping based on allelotyping of SNPs and assembly of three-dimensional display of allelic losses on chromosome 17p13 in a single cystectomy specimen. **(a)** Genome sequence map of a 4.27-Mb region on chromosome 17p13 (www.mdanderson.org/departments/genomemaps/). The positions of hypervariable DNA markers and positions of the known or predicted genes are depicted by the bars on the left side of the map. The bars on the right side of the map designate the positions of all informative (polymorphic) SNPs identified in the peripheral blood of the same patient. Monomorphic (noninformative) SNPs are not shown. **(b)** Map of the deletions designated by vertical red bars. Their predicted borders were defined by the nearest flanking SNPs, which retained polymorphism in all mucosal samples (see Figure 4). The bars on the right side designate the positions of SNPs with allelic loss. Their distribution patterns in bladder mucosa are depicted by the shaded blocks in c. **(c)** Three-dimensional display of allelic losses depicted by shaded blocks in relation to the distribution of transitional cell carcinoma (TCC) and their association with *in situ* preneoplastic lesions. **(d)** Graphic representation of histologic map of cystectomy specimen with the distribution of *in situ* preneoplastic lesions that progressed to invasive TCC. Histologic map code is identical to the code shown in Figure 1. **(e)** Summary of binomial maximum likelihood analysis testing the relationship among loss of SNP polymorphism and progression of urothelial neoplasia from *in situ* precursor conditions to invasive TCC. Cumulative LOD scores for SNPs with loss of polymorphism were calculated at variable $\theta = (0.01, 0.5, \text{ and } 0.99)$ and tested against T_{\max} . The significance of clonal losses was analyzed for normal urothelium (NU), low-grade intraurothelial neoplasia (LGIN), high-grade intraurothelial neoplasia (GHIN), and TCC. To simplify the data, stringency 1 calculations are presented only. The patterns of significant LOD scores are as described in Methods. Note that the significant pattern of LOD scores typically paralleled high T_{\max} values (○, LOD score < 3 ; ●, LOD score ≥ 3). A cluster of two discontinuous segments of allelic loss associated with clonal expansion, which form overlapping large plaques, not only involving microscopically recognizable *in situ* preneoplastic lesions but also extending into areas of microscopically normal urothelium of bladder mucosa was identified. The segments were defined by clusters of neighboring SNP with the loss of polymorphism as shown by brackets on the right side of the three-dimensional diagram.



Results

For mapping with SNPs, the sequence-based maps spanning approximately 27 and 5 Mb centered around RB1 and p53, respectively, were assembled. The integrated gene and SNP maps of the regions were used to select 661 and 960 SNPs, which were genotyped by pyrosequencing. Genotyping of SNPs was performed in three sequential steps on DNA

samples corresponding to histologic maps of the entire bladder mucosa in human cystectomy specimens with invasive urothelial carcinoma. Initially, all selected SNPs of normal genomic DNA were sequenced. In the next step, those SNPs that exhibited polymorphism were tested on representative DNA samples corresponding to invasive transitional cell carcinoma of the bladder (TCC). In the final step, those SNPs that showed allelic loss were



tested on all mucosal samples of the same cystectomy specimen. The distribution of allelic loss for each SNP was subsequently superimposed over the histologic map of the entire organ. The borders of deleted segments were defined by the nearest flanking SNPs that retained polymorphism in all mucosal samples. The allelic loss was determined by the comparison of the sequences of the mucosal and tumor DNA samples to the sequences of nontumor DNA from the same patient. Clonal loss was defined as the loss of the same polymorphic nucleotide in multiple mucosal samples. Nonclonal loss was defined as the loss of signal interchangeably corresponding to both residues of the polymorphic SNP in different mucosal samples.

The initial testing of 661 and 960 SNPs of peripheral blood samples selected for mapping of 13q14 and 17p13 regions identified 101 and 187 polymorphic SNPs mapping to 27- and 4-Mb segments containing RB1 and p53, respectively. Subsequent testing of the informative polymorphic SNPs in representative DNA samples corresponding to invasive TCC identified 15 and 11 haplotypic SNPs defining allelic loss in 13q14 and 17p13, respectively. These SNPs were tested on all mucosal samples of two cystectomy specimens. By superimposing the distribution of allelic losses over histologic maps of the entire bladder mucosa, we identified two basic distribution patterns of losses: scattered and plaque-like (Figure 1). The scattered foci of allelic losses in individual samples of bladder mucosa were typically nonclonal and involved both polymorphic residues of the SNP. In contrast, plaque-like losses were clonal, involving large areas of bladder mucosa with various precancerous conditions and typically extending to normal appearing urothelium. Such large clonal plaques of allelic loss were statistically significantly associated with progression of *in situ* preneoplasia to invasive TCC as revealed by binomial likelihood analysis. This association implied that the loss was responsible for growth advantage, which resulted in *in situ* expansion of a preneoplastic clone. The deleted regions associated with clonal expansion spanned from 0.001 to 4.3 Mb (average 0.67 Mb) and formed clusters of discontinuously deleted segments separated by nondeleted regions (Figures 2 and 3). In some instances, the flanking SNPs disclosed allelic

loss restricted to an individual mucosal sample, a reflection of the instability of regions contiguous to deleted segments (Figure 4).

A cluster of deletions associated with an *in situ* clonal expansion in 13q14 involved approximately 13.97 Mb and consisted of four deleted segments spanning 0.79, 0.68, 1.91, and 4.68 Mb. These segments contained 7, 8, 9, and 10 positional candidate genes, respectively, which might lead to a growth advantage for a preneoplastic clone. The model tumor suppressor gene RB1 mapped to the deleted segment between 46.35 and 47.03 Mb (Figure 5). A similar cluster of deletions that consisted of two deleted segments spanning 0.12 and 0.05 Mb was identified 0.8 Mb telomerically to p53 in the 17p13 region. These two segments contained 12 and five positional candidate genes, respectively, which might also be relevant for growth advantage of a preneoplastic clone.

Visually, the segments of deletions associated with growth advantage and *in situ* clonal expansion of preneoplastic urothelial cells contained a significantly higher concentration of repetitive sequences as compared to the nondeleted segments and deletions, which formed scattered nonclonal random mucosal foci (Figures 5 and 6). The statistical analysis confirmed this impression and disclosed that for chromosome 17 the segments of allelic loss associated with a plaque-like clonal expansion of preneoplastic clone involving large areas of bladder mucosa contained a significantly higher concentration of non-LTR repeats as compared to the regions with retention of polymorphism and regions, which showed nonclonal loss in the form of scattered independent foci (Table 1 and Figure 7). No such relationship could be found for allelic loss on chromosome 13. Similar analyses of CpG island and gene content did not reveal any relationship between their densities and positions of deleted segments (data not shown).

Discussion

Whole-organ histologic and genetic mapping facilitated the analysis of the relationship between allelic loss and disease development from morphologically normal urothelium through microscopically recog-

Figure 4 Clusters of clonal and nonclonal loss and their association with progression of bladder neoplasia. (a) SNPs pyrograms showing nucleotide incorporation signals of SNPs in peripheral blood and corresponding mucosal samples. Pyrograms 1, 4, 5, and 7 correspond to flanking SNPs, which show retention of polymorphism in all mucosal samples. Pyrograms 2 and 6 show the loss of the same allele in multiple mucosal samples that form a continuous plaque involving a large area of bladder mucosa, while pyrogram 3 shows nonclonal loss of SNP alleles in scattered mucosal areas. (b) Two discontinuous segments of deleted regions on chromosome 17 are depicted by shadowed areas spanning 0.12 and 0.05 Mb, respectively. (c) The positions of all known and predicted genes within and around deleted regions are depicted by bars on the left side of the map. The positions of the informative SNPs are designated by the bars on the right side of the map. The sites in the bladder mucosa that were the sources of SNPs showing loss of polymorphism are connected to their respective areas depicted in d. (d) Histologic map of cystectomy specimen. Areas outlined by the solid and dashed lines show clonal loss of their corresponding SNPs, which formed a large plaque in the bladder mucosa. Areas outlined by the dotted line show nonclonal loss of polymorphism in separate mucosal areas. (e) Representative microscopic samples of NU, precursor *in situ* conditions, and TCC are shown. Note the increasing nuclear atypia and architectural disorder as intraurothelial precursor conditions advanced from LGIN to HGIN and invasive stromal growth in TCC. Solid black bars within photomicrographs indicate 50 μ m.

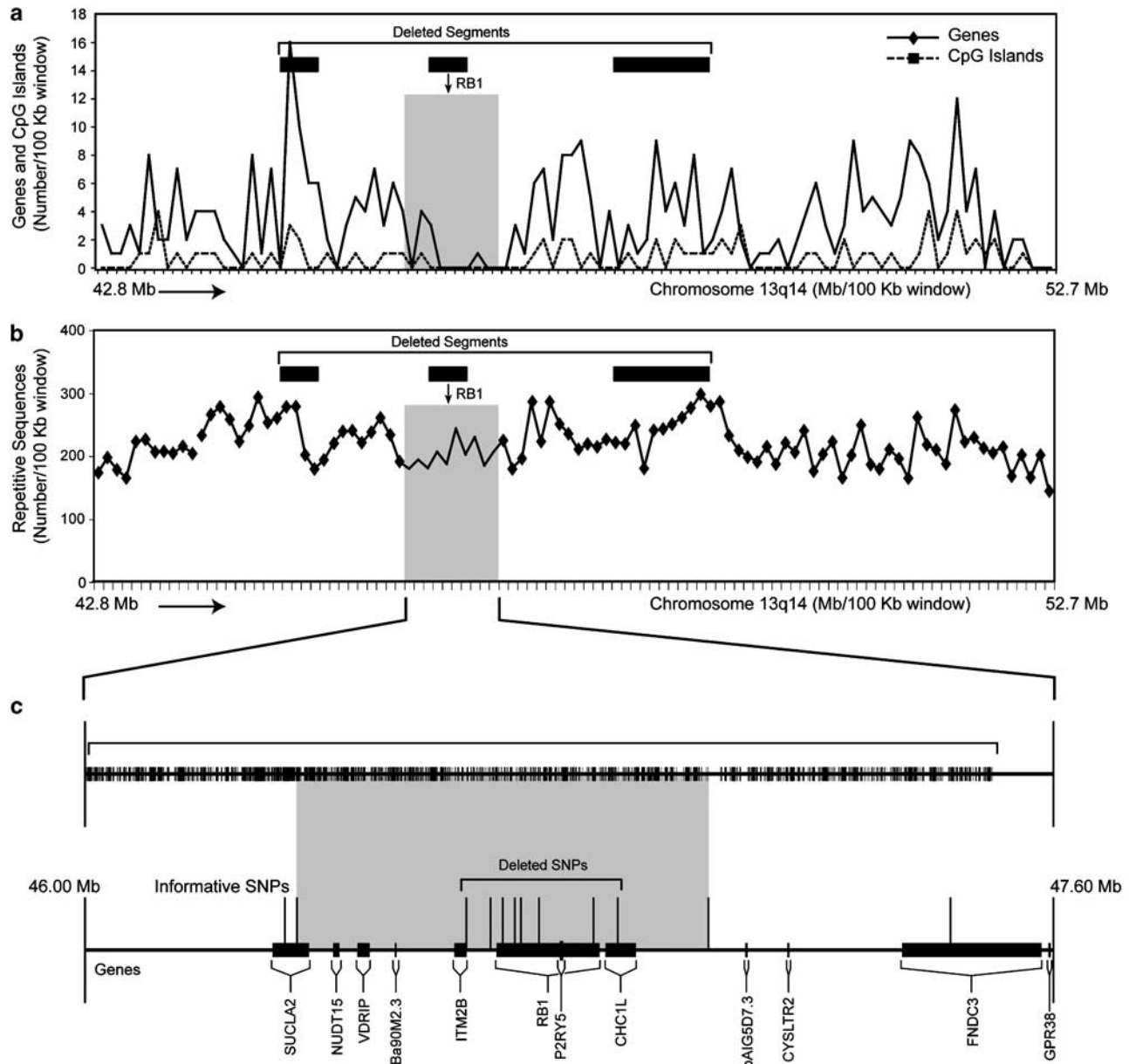


Figure 5 Genomic organization and deletions of 10-Mb region around RB1. (a) Distribution of CpG islands and genes in a 100-kb window. (b) Distribution of all non-LTR repetitive sequences in a 100-kb window: (c) Expanded view of the 1.6-Mb region around RB1 showing the positions of repetitive sequences and genes. Black bars in (a) and (b) depict the positions of three deleted segments associated with *in situ* clonal expansion of urothelial cells identified by SNP-based whole-organ mapping shown in Figure 2. Shaded areas in (c) and (d) depict a deleted segment containing eight genes including RB1. Note that the deleted segments shown in (a) and (b) overlap with regions containing high concentrations of repetitive sequences.

nizable intraurothelial lesions to invasive disease in the entire mucosa of the resected organ. We precisely identified those losses that are associated with the *in situ* clonal expansion and are likely significant for the incipient phases of bladder carcinogenesis. Specifically, we identified clusters of discontinuous loss mapping to 13q14 and 17p13 containing model suppressor genes RB1 and p53. The high density of SNP markers permitted us to narrow the deleted segments from 4.6 Mb to below 0.1 Mb scale and disclosed that losses repre-

sented clusters of discontinuous segments ranging in size from 0.1 to 4.6 Mb separated by nondeleted regions.

The strategy presented here is not only important for identifying positional candidate genes that may provide growth advantage for a preneoplastic clone, but for also permitting the analysis of the overall genomic landscape of the involved regions. Such an analysis disclosed that the deleted regions associated with clonal expansion of preneoplasia contained a high concentration of repetitive

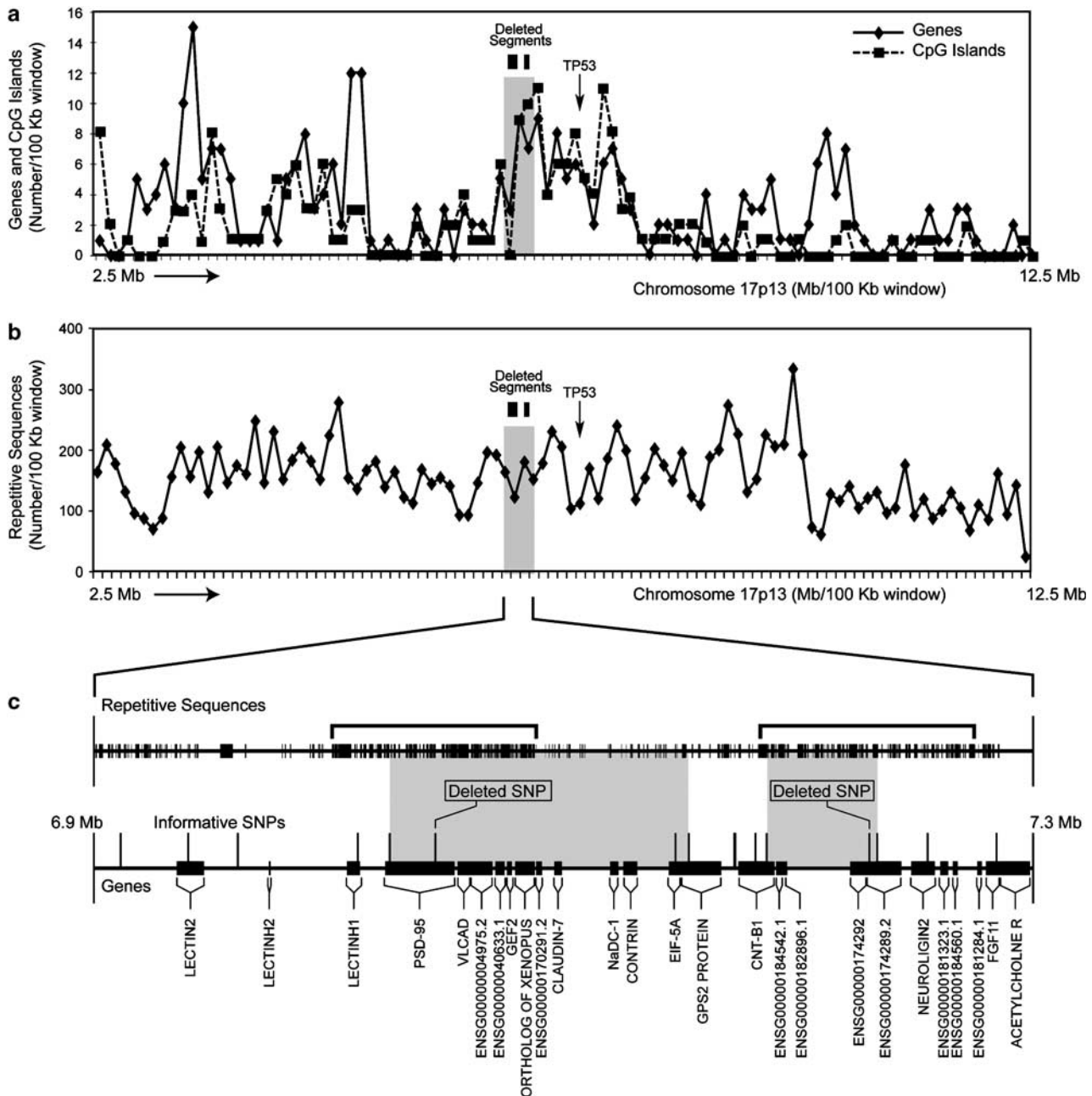


Figure 6 Genomic organization and deletions of 10-Mb region around TP53. **(a)** CpG islands and genes in a 100-kb window. **(b)** All non-LTR repetitive sequences in a 100-kb window. **(c)** Expanded view of the 0.4 Mb region located telomerically to TP53 showing the positions of repetitive sequences and genes. Black bars in **(a)** and **(b)** depict the positions of two deleted segments associated with *in situ* clonal expansion of urothelial cells identified by SNP-based whole-organ mapping shown in Figure 3. Shaded areas in **(c)** and **(d)** depict two deleted segments containing 12 and 5 genes, respectively. Note that the deleted segments are associated with high concentrations of repetitive sequences, marked in **(c)** by the brackets.

sequences. This, in turn, suggests that the unique composition of noncoding DNA in the regions may predispose the genome to instability and allelic imbalance. The evidence provided by our study implies that the genes located in deleted regions may be responsible for growth advantage of *in situ* preneoplasia and the high content of repetitive sequences in these regions predisposed to deletions

is circumstantial and has to be elucidated by functional studies. In summary, the strategy presented in this communication permits precise identification of deleted chromosomal regions associated with clonal expansion of preneoplasia. Such maps performed on a whole-genome scale will be critical for future specific studies on mechanisms initiating carcinogenesis.

Table 1 Analysis of non-LTR repeat densities in deleted and nondeleted segments of 13q14 and 17p13

	Repeats densities around SNPs ^a		
	Retention of polymorphism	Scattered nonclonal loss	Plaque-like clonal loss
Chromosome 13q14	10 051 ± 828 ^b	8235 ± 959	9367 ± 650
	$P < 0.5313^*$		$P < 0.4337$
	$P < 0.6447$		
Chromosome 17p13	10 380 ± 252	8906 ± 941	18 577 ± 3100
	$P < 0.1639$		$P < 0.003$
	$P < 0.0003$		

^aThe repeat sequences densities were defined as the number of repeat base pairs in a fixed width window as described in the Methods.

^bMean value of non-LTR repeat density ± standard error of the mean.

*P-value by Wilcoxon's test for paired samples.

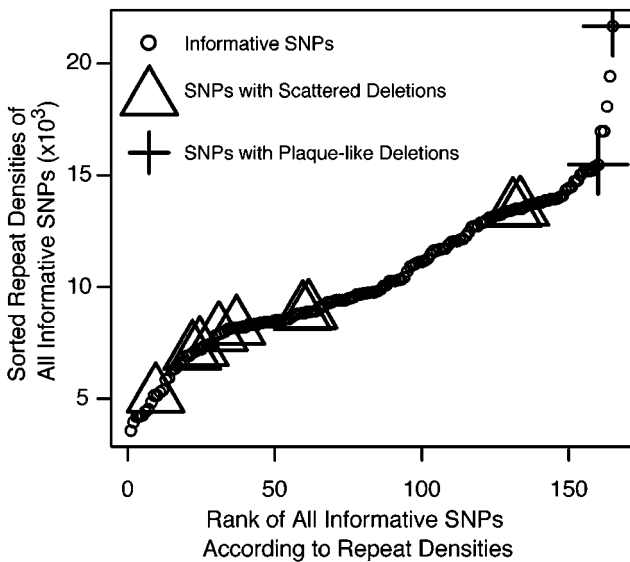


Figure 7 The analysis of non-LTR repeat densities within 5Mb segment around p53 on chromosome 17p13. The map of the analyzed segment with positions of the genes and SNPs is shown in Figure 3. Note that SNPs with a plaque-like clonal expansion of allelic loss are surrounded by a high concentration of non-LTR repeats.

Acknowledgements

This work was supported by the National Institute of Health, UO1 CA85078 (BC), RO1 CA066723 (BC), and P50 CA91846 (Pr. 1, BC).

References

- Albores-Saavedra J. The Pathology of Incipient Neoplasia. Oxford University Press: New York, 2001.
- Dinney CP, McConkey DJ, Millikan RE, *et al*. Focus on bladder cancer. *Cancer Cell* 2004;6:111–116.

- Izawa JI, Slaton JW, Kedar D, *et al*. Differential expression of progression-related genes in the evolution of superficial to invasive transitional cell carcinoma of the bladder. *Oncol Rep* 2001;8: 9–15.
- Carethers JM. The cellular and molecular pathogenesis of colorectal cancer. *Gastroenterol Clin North Am* 1996;25:737–754.
- Ilyas M, Straub J, Tomlinson IP, *et al*. Genetic pathways in colorectal and other cancers. *Eur J Cancer* 1999;35: 1986–2002.
- Czerniak B, Li L, Chaturvedi V, *et al*. Genetic modeling of human urinary bladder carcinogenesis. *Genes Chromosomes Cancer* 2000;27:392–402.
- Houlston RS, Peto J. The search for low-penetrance cancer susceptibility alleles. *Oncogene* 2004;23: 6471–6476.
- Czerniak B, Chaturvedi V, Li L, *et al*. Superimposed histologic and genetic mapping of chromosome 9 in progression of human urinary bladder neoplasia: implications for a genetic model of multistep urothelial carcinogenesis and early detection of urinary bladder cancer. *Oncogene* 1999;18:1185–1196.
- Yoon DS, Li L, Zhang RD, *et al*. Genetic mapping and DNA sequence-based analysis of deleted regions on chromosome 16 involved in progression of bladder cancer from occult preneoplastic conditions to invasive disease. *Oncogene* 2001;20:5005–5014.
- Kram A, Li L, Zhang RD, *et al*. Mapping and genome sequence analysis of chromosome 5 regions involved in bladder cancer progression. *Lab Invest* 2001;81: 1039–1048.
- Chaturvedi V, Li L, Hodges S, *et al*. Superimposed histologic and genetic mapping of chromosome 17 alterations in human urinary bladder neoplasia. *Oncogene* 1997;14:2059–2070.
- Tong BC, Dhir K, Ha PK, *et al*. Use of single nucleotide polymorphism arrays to identify a novel region of loss on chromosome 6q in squamous cell carcinomas of the oral cavity. *Head Neck* 2004;26:345–352.
- Zhou X, Li C, Mok SC, *et al*. Whole genome loss of heterozygosity profiling on oral squamous cell carcinoma by high-density single nucleotide polymorphic allele (SNP) array. *Cancer Genet Cytogenet* 2004; 151:82–84.

- 14 Janne PA, Li C, Zhao X, *et al*. High-resolution single-nucleotide polymorphism array and clustering analysis of loss of heterozygosity in human lung cancer cell lines. *Oncogene* 2004;23:2716–2726.
- 15 Mostofi F. Histological Typing of Urinary Bladder Tumors, 2nd edn. Springer: Berlin, New York, 1999.
- 16 Sobin L, Wittekind C. TNM Classification of Malignant Tumors. John Wiley: New York, 1997.
- 17 Ronaghi M. Pyrosequencing for SNP genotyping. *Methods Mol Biol* 2003;212:189–195.
- 18 Ahmadian A, Gharizadeh B, Gustafsson AC, *et al*. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem* 2000;280:103–110.
- 19 Hartigan J. Clustering Algorithms. John Wiley & Sons, Inc.: New York, 1975.
- 20 Ott J. Analysis of Human Genetic Linkage. Johns Hopkins University Press: Baltimore, 1991.